Examiners' Report
Principal Examiner Feedback

Summer 2017

Pearson Edexcel GCE Mathematics

Statistics S2 (6684)

Edexcel and BTEC Qualifications

Edexcel and BTEC qualifications are awarded by Pearson, the UK's largest awarding body. We provide a wide range of qualifications including academic, vocational, occupational and specific programmes for employers. For further information visit our qualifications websites at www.edexcel.com or www.btec.co.uk. Alternatively, you can get in touch with us using the details on our contact us page at www.edexcel.com/contactus.

Pearson: helping people progress, everywhere

Pearson aspires to be the world's leading learning company. Our aim is to help everyone progress in their lives through education. We believe in every kind of learning, for all kinds of people, wherever they are in the world. We've been involved in education for over 150 years, and by working across 70 countries, in 100 languages, we have built an international reputation for our commitment to high standards and raising achievement through innovation in education. Find out more about how we can help you and your students at: www.pearson.com/uk

Summer 2017
Publications Code 6684_01_1706_ER

Statistics 2 (6684) – Principal Examiner's report

General introduction

On the whole this paper was well answered. Most questions on the paper seemed accessible to the students. Generally the work was quite well presented.

The paper proved accessible to the majority of students and there was little evidence of there not being enough time to complete the paper. There were the usual arithmetic and algebraic errors, but it was clear many students were out of practice in techniques, taught in Statistics 1 such as using an inverse normal distribution and using conditional probability.

Question 1

In part (a) the vast majority of students could identify the correct binomial distribution for the problem and therefore make a start on the question. However, the most common reason for lost marks was through writing incorrect (usually $P(X \leq 9) = 0.9827$) or even no probabilities from the relevant binomial which was clearly asked for in the question. Students should take note, and read the requirements of the question carefully before proceeding.

By far the most common error in terms of the critical region was students selecting the upper region to be $X \geq 8$ and occasionally $X \leq 2$ was seen for the lower critical region. There are still a few students who are writing their critical regions as $P(X \leq 1)$ or $P(X \geq 9)$ rather than in the required form of $X \leq 1$ or $X \geq 9$

The majority of students were able to provide a fully correct solution to part (b). Students seem to understand the relevant information which needs to be provided when concluding a hypothesis test to gain credit. Common errors were writing the wrong or incomplete hypotheses such as $H_0 = 0.2$ or using words rather than symbols and occasionally the wrong distribution was used. Most students opted for comparing relevant percentages rather than finding the critical region in this question and did so correctly. The most common reason for lost marks was using $P(X < 6)$ rather than $P(X \leq 6)$ for their test statistic. It was pleasing to see that many of the students are now correctly writing their conclusions in context using the key words from the question.

Question 2

Part (a)(i) was answered extremely well overall with virtually all of the students gaining full marks. However, a few students attempted to find $P(X > 4)$ or $P(X = 4)$ in error. The vast majority of students also earned full marks in part (a)(ii), though more errors were seen than in part (a)(i). Some students used an incorrect value for λ, usually 2.5, having not realised that the new mean was 0.625. However, even when an incorrect value for λ was used, the correct method for finding $P(X = 3)$ was generally applied.

In complete contrast to part (a), very few students gained any marks in part (b). This part of the question proved to be extremely challenging to all but the most able students, with many clearly not understanding the meaning of the question and hence being unable to access it at all. Of those that did show some semblance of knowing how to approach this question, by far the most common error was to associate $P(X = 0)$ with 0.2, usually $P(X = 0) < 0.2$ or $P(X = 0) = 0.2$ and sometimes $P(X = 0) > 0.2$. These students would then either be unable to proceed any further or continued with $e^{-\lambda} < 0.2$, $e^{-\lambda} = 0.2$ or $e^{-\lambda} > 0.2$, accordingly.

Most students who began with $P(X \geq 1) < 0.2$ or $P(X = 0) > 0.8$, did ultimately achieve the correct answer, even when students correctly obtained $e^{-\lambda} > 0.8$, some did not use an appropriate λ, some forgot to convert to minutes and in a few instances, an answer of 5.36 was incorrectly rounded up to 6 minutes.

Part (c) was answered well overall, with many students gaining full marks. Hypotheses were generally stated clearly and correctly, with a mean rate of 2.5, or more commonly 5. Occasionally, p (or no letter) was used incorrectly in place of λ or μ, or 10 used as the mean rate. Other errors included, finding $P(X \leq 10)$; $1 - P(X \leq 10)$; $P(X = 10)$; using Po(2.5) with $P(X \geq 5)$; and stating the critical region as $X > 9$ or just 10. Attempting to identify the critical region was generally less successful as an approach. Correct non-contextual statements tended to be obtained, although occasionally students included contradictory non-contextual statements or incorrectly interpreted (or made incorrect) comparisons with 0.05, 0.95 or 10

Whilst, on occasion, students neglected to place these conclusions into context, on the whole students were able to draw the correct contextual conclusions. However, in a few instances, these conclusions were not fully contextual and did not refer to calls.

Question 3

Part (a) was well answered by the vast majority of students who were able to gain full credit and despite the odd numerical error most gained at least the method marks by applying the given formula. A common mistake was to simply integrate $f(x)$ or to just write $xf(x)$ without going on to integrate or to use an incorrect lower limit of 0 rather than 1.

Part (b) was often answered well with most students using the most efficient method of integrating between the correct limits 2.5 and 4 correctly. Some students opted to find the cumulative distribution function, with less success. Often these students gained little or no credit, especially if they failed to show evidence of using correct limits, with many neglecting to use 1 as the lower limit, presumably assuming it was 0. A minority of students found the answer to be $\frac{5}{8}$ and forgot to subtract it from 1.

In part (c), most students realised they simply needed to square the answer to part (b) and often gained the method mark even if this was the first mark gained in the entire question. Occasionally students needlessly wrote down the full binomial distribution and used the relevant formula but still gained full credit. However a small number of students doubled rather than squared their answer to b.

Part (d) proved to be a good challenge for students and a range of responses were seen. Many could correctly identify 0.768 for the first B mark (usually those who answered (b) correctly). From there, some students seemingly got confused as to exactly what the question was asking, and a variety of responses based on previously calculated values were seen. By far the most common mistake was to use their (b) on the numerator of the fraction rather than their (c). Compounding this error was some multiplication of probabilities rather than division.

Question 4

For many students this was good source of marks with many correct solutions. In part (a) the majority of students found the equation from the given E(X) of the continuous uniform distribution but a significant number of students failed to find two correct equations. The most usual mistake in this stage was to write $\frac{5}{\beta-\alpha} = \frac{3}{5}$. Many students who did have two correct equations made simple sign errors or failed to use brackets correctly.

There was a real lack of drawing an accurate sketch of the pdf; while this was not strictly required, it would have been useful and there seemed to be a correlation between those who did this and those who arrived at a correct solution.

In part (b), those students who gained full marks for part (a) had no problems in deducing the value of c. However, incorrect answers to part (a) really caused problems for students attempting part (b). They spent a lot of time wrestling with awkward equations using unpleasant values from (a).

Most students used $\frac{1}{\beta-\alpha} \times (9 - c)$ but some did use F(9) – F(c), thus they were able to gain the M mark on this question irrespective of whether they found the correct value for c. Amongst students who had made errors in early parts of the question it was frequently necessary to consider the special case highlighted in the mark scheme. These students had usually recognised that the interval from c to 9 did not wholly lie within their distribution and had taken this into account.

Nearly all students attempted part (c), regardless of their success in the earlier parts. However, many did not consider the separate cases stemming from the two dimensions of the rectangle and gave 0.3 or 0.5as their answer.

Question 5

In part (a), most students showed a good knowledge of how to standardise. However, surprisingly few were able to complete part (a) accurately and hence be awarded all 3 marks. The most common error was using a z-value not from the percentage points table, and hence lost the B mark due to their value being only 3.s.f.

Other errors were: adding an unnecessary continuity correction, using the variance instead of the standard deviation in their standardisation, equating to a probability rather than a z-value, not using the correct sign of the z-value and sometimes ending up on a negative standard deviation, and lastly, leaving their answer as an exact fraction.

Part (b) was well answered with the vast majority of students scoring full marks. Even if they could not find the correct model most managed to gain the method mark by writing $P(X < 2)$.

Part (c) was a challenge, with very few students scoring full marks. Most students correctly identified the mean and variance for the normal approximations but many did not use the correct continuity correction

Most students managed the correct z-value but a number used the wrong sign in their equation. Again, some students incorrectly used a probability as their z-value.

The majority of students were able to form a quadratic of some sort from their standardisation equation, and many were able to attain the M mark for a correct method of solving this. However, far too many simply produced solutions to their quadratic from no working and hence lost marks if they had used an incorrect equation.

Clearly having a quadratic in $\sqrt{n}$ was challenging, and many students had no method to deal with this – often failing to identify how it mapped to a "standard" quadratic. Some students showed good initiative and exhibited useful methods such as squaring both sides or using a substitution.

Question 6

A lot of marks were generally lost on this question from part (c) onwards. Most students correctly drew a trapezium shaped pdf in part (a) and labelled the correct values on the coordinate axis. The occasional missing value was seen notably 'k' or '$\frac{1}{3}$'. Common incorrect answers included a 'peak' shaped graph with a maximum value at $x = 4$. Occasionally curves were seen, gaining no credit.

A large number of students chose to inefficiently use integration in part (b) albeit often correctly. When a question offers 2 marks, it is clear the method required should not take extended working. Working out the area of a trapezium was the most efficient method, but splitting the area into two triangles and a rectangle was often seen and done correctly.

A fairly standard technique was required in part (c) so it was surprising to see it so poorly completed on occasions. Limits were the main problem in this part. Several students just put the upper limit as x with no attempt to add in F(3) or F(5). The B mark was sometimes the only mark awarded, and it was pleasing to see the vast majority of students were able to gain this mark.

Many students went on to gain the method marks in part(d) by equating their cdf to 0.9 but didn't always go onto solve it. A sizeable number of students seem to be very quick to solve a quadratic on their calculator showing no method to get the answer from their incorrect quadratic, meaning the method mark for solving a quadratic equation could not be awarded. A small number of students did not appreciate the significance of the two answers and omitted to disregard 6.77 in their final answer.

In part (e), once students noted that, due to symmetry E(X) = 4, they would often proceed to the correct answer quickly, however there were a number of students who failed to spot this with a minority trying to work out the mean by integration, rarely getting to E(X) = 4.

Pearson Education Limited. Registered company number 872828
with its registered office at 80 Strand, London, WC2R 0RL, United Kingdom